

Analisis Sentimen Masyarakat Pada Komentar Produk Eiger Di Instagram Menggunakan Metode Naive Bayes Classifier

Robbol Hidayatullah¹, Ulya Anisatur Rosyidah¹, Ginanjar Abdurrahman¹

Universitas Muhammadiyah Jember

e-mail : robbohdt@gmail.com, ulyaanisatur@unmuhjember.ac.id, abdurrahmanganjar@unmuhjember.ac.id

DOI: <https://doi.org/10.32528/nms.v1i6.232>

*Correspondensi: Robbol Hidayatullah

Email: robbohdt@gmail.com

Published: November, 2022



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstrak: Eiger atau PT Eigerindo Multi Produk Industri merupakan perusahaan yang memproduksi pakaian dan perlengkapan hiburan alam atau outdoor. Eiger dalam pemasarannya juga menggunakan media sosial yaitu Instagram. Dalam setiap produk yang diunggah pada laman akun Instagram Eiger yaitu @eigerindostore mendapatkan berbagai komentar dari masyarakat. Dalam hal ini penelitian memanfaatkan data komentar tersebut sebagai data penelitian ini yaitu sentimen analisis. Pada penelitian ini metode yang digunakan sebagai klasifikasi adalah Naive Bayes sedangkan metode yang digunakan dalam skenario pengujian adalah split data dengan data latih 80% dan data uji 20% dan skenario Cross Fold Validation dengan nilai k yaitu 2, 4, 5, 7 dan 10. Dari hasil implementasi metode Naive Bayes terhadap data komentar pada akun Instagram Eiger diperoleh hasil yaitu rata-rata akurasi yang diperoleh yaitu 85.27%, rata-rata presisi yang diperoleh adalah 88.29% dan rata-rata recall yang diperoleh adalah 81.26%. Nilai akurasi tertinggi diperoleh pada k = 10 dengan fold ke 10 yaitu 91.07%. Nilai presisi tertinggi diperoleh pada k = 10 fold ke 10 yaitu 90% dan nilai recall tertinggi diperoleh pada k = 7 fold ke 7 yaitu 91.67%

Keyword: Analisis Sentimen, Cross Fold Validation, Eiger, Klasifikasi, Naive Bayes.

PENDAHULUAN

Eiger atau PT Eigerindo Multi Produk Industri adalah perusahaan dan merek dari Indonesia yang memproduksi pakaian dan perlengkapan hiburan alam. Perusahaan memproduksi produk seperti tas dan jaket yang digunakan dalam kegiatan mendaki gunung, berkemah dan panjat tebing.

Fitur komen pada Instagram sering digunakan oleh masyarakat untuk memberi nilai, kepuasan, emosi pada produk yang mereka beli dari toko produk tersebut. Misalnya, dari akun Instagram @eigerindostore dengan produk sepatu gunung tipe Bugle Shoes mendapat komentar penilaian negatif dari salah satu pelanggan atas nama @solihin.bahri dengan komentar “Pertama dipakai ke rakum putih bahan di atas vibram lapisan putih ndk begitu kuat min..pulangnyanya lemnya kelihatan sbagian terbuka”. Tidak hanya komentar negatif mengenai produk Eiger, komentar positif juga diperoleh, seperti komentar yang ditulis oleh akun @lolitaasiswanto pada produk jaket gunung Kahama yang berisi “Kahama alhamdulillah udah punya, Min. Ngga mengecewakan”. Komentar-komentar pada produk Eiger di Instagram ini yang memiliki nilai sentimen dapat digunakan menjadi data yang selanjutnya dapat digunakan oleh perusahaan untuk perbaikan produk ke depannya. Hal tersebut dapat dilakukan dengan cara teknik analisis sentimen.

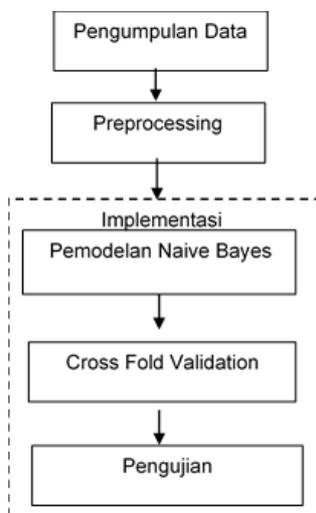
Teknik untuk menentukan klasifikasi positif atau negatif salah satunya adalah menggunakan teknik Machine Learning. Ada banyak teknik klasifikasi yang dapat digunakan dalam Machine Learning, salah satunya Naive Bayes. Naive Bayes Classifier merupakan metode sederhana berdasarkan probabilitas, tetapi memiliki akurasi dan performa yang tinggi dalam mengklasifikasikan teks.

Eiger atau PT Eigerindo Multi Produk Industri adalah sebuah perusahaan dari Indonesia yang memproduksi pakaian dan perlengkapan hiburan alam. Perusahaan memproduksi produk seperti tas dan jaket yang digunakan dalam kegiatan mendaki gunung, berkemah dan panjat tebing. Eiger juga menggunakan media sosial dalam pemasaran atau promosinya. Salah satunya adalah Instagram, pada

media sosial ini diunggah beberapa produk Eiger. Dalam unggahan setiap produk Eiger juga mencantumkan diskon dan spesifikasi produk yang dipasarkan.

METODE

A. Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Sumber : JITTER 2015

HASIL DAN PEMBAHASAN

A. Gambaran Data

Data pada penelitian ini diperoleh dari hasil scraper atau mengunduh data komentar pada laman akun Instagram @eigerindostore. Scraping dilakukan pada 107 unggahan produk Eiger pada Instagram @eigerindostore menggunakan extension atau plugin pihak ketiga pada Google Chrome yaitu IGCommentExport dari WeBooster. Dari 107 produk diperoleh data komentar sebanyak 5688 data komentar.

B. Validasi Sentimen Data Komentar

Tahap validasi pada penelitian ini yaitu membagi data komentar menjadi tiga kategori yaitu positif, negatif dan netral. Data yang akan digunakan pada penelitian ini adalah data positif dan negatif.

Total data pada penelitian ini adalah 5688 dengan pembagian data yang tersaji pada gambar 4.2 yaitu 4988 data netral, 439 data positif dan 261 data negatif. Data ini hanya akan digunakan data sentimen positif dan negatif sehingga 4988 data netral akan dihapus dan tersisa 700 data yang akan dianalisis.

C. Cleansing

Data pada kolom Comment Text masih berbentuk raw atau mentah. Di dalamnya terdapat simbol-simbol, emoticon atau pun link tautan. Hal itu dinilai mengganggu dalam analisis sentimen nantinya dan pada proses ini hal tersebut akan dibersihkan atau dihapus.

D. Tokenizing

Text preprocessing kedua adalah Tokenizing atau memecah kalimat menjadi per kata. Hal ini bertujuan agar tiap kata dapat dianalisis dan jika menemukan kata yang tidak perlu dapat dihapus. Tokenizing juga bertujuan dalam seleksi kata pada proses stemming nantinya.

E. Stopword Removal

Proses selanjutnya adalah Stopword Removal atau menghapus kata yang keberadaannya tidak berpengaruh pada sebuah kalimat sentimen. Pada proses ini Stopword list disediakan oleh package Nltk dalam bentuk kata baku. Untuk mengantisipasi kata tidak baku, penulis menambahkan sendiri secara manual.

F. Stemming

Proses terakhir pada text preprocessing adalah Stemming atau mengembalikan kata yang memiliki awalan atau imbuhan ke kata dasar awalnya. Pada proses ini penulis juga memanfaatkan package Sastrawi untuk melakukan Stemming.

G. Pembobotan TF-IDF

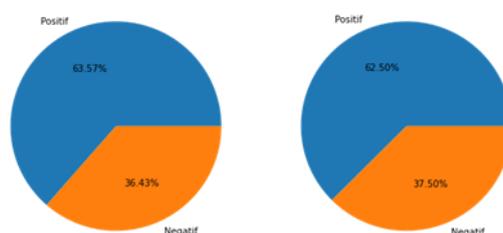
Data hasil proses text preprocessing akan dihitung bobot tiap katanya. Pada proses ini digunakan pembobotan TF-IDF, merujuk pada bab 2 pada sub bab TF-IDF berikut persamaan yang digunakan untuk menghitung pembobotan pada kata :

$$IDF = tf_{ij} \times \log \left(\frac{D}{df_{ij}} \right).$$

H. Implementasi Naive Bayes, Cross Fold Validation, Split Data dan Pengukuran

Hasil pada proses sebelumnya adalah tiap kata memiliki bobot tersendiri. Pada proses ini akan dilakukan implementasi Naive Bayes untuk membentuk model klasifikasi dan Cross Fold Validation untuk mengetahui lipatan data mana yang memiliki akurasi tertinggi. Pada penelitian ini juga diimplementasikan metode Split Data sebagai temuan lain dalam penelitian ini bertujuan sebagai pembandingan dari hasil pengukuran pada skenario uji K Fold Cross Validation. Skenario pada Split Data yang digunakan adalah 80% data latih dan 20% data uji. Pada penelitian ini nilai k pada Cross Fold Validation yang digunakan adalah 2, 4, 5, 7 dan 10 serta split 20% data uji 80 data latih.

Uji coba pertama adalah split data, dimana pada skenario ini total 700 data dibagi menjadi data latih 80% dan data uji 20%.

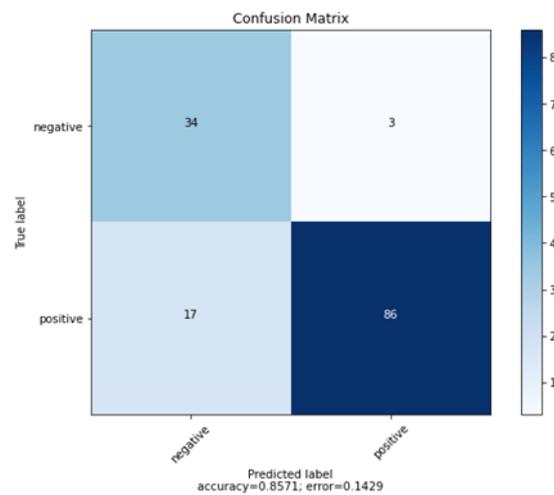


Gambar 2. (kiri) Pembagian data latih Split data. (kanan) pembagian data uji Split data

Sumber : Hasil Perhitungan

Dari skema Split data ini diperoleh hasil pada 80% data latih adalah 560 data sedangkan pada data uji berjumlah 140 data. Dari 560 data latih diperoleh 350 data bernilai sentimen positif dan 210 data bernilai sentimen negatif. Dari 140 data uji diperoleh 89 data bernilai sentimen positif dan 51 data bernilai sentimen negatif.

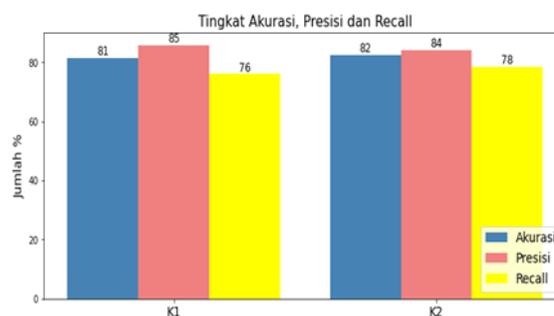
Hasil klasifikasi dilakukan perbandingan pada data aktual untuk menentukan nilai True Positive, True Negative, False Positif dan False Negative dalam bentuk Confusion Matrix. Berikut hasil Confusion Matrix digambarkan pada gambar di bawah ini.



Gambar 3. Confusion matrix hasil klasifikasi Split data
Sumber : Hasil Perhitungan

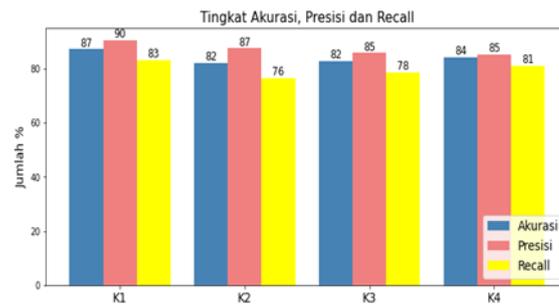
Dari gambar di atas dapat dijelaskan bahwa kriteria True Positive memperoleh nilai 86 data, kriteria True Negative memperoleh 34 data, kriteria False Positive memperoleh nilai 3 data dan kriteria False Negative memperoleh nilai 17 data.

K = 2, Pada skenario ini yang dilakukan 2 kali pengujian dan diperoleh hasil pengukuran tingkat akurasi yaitu 81.43% pada pengujian pertama dan 82.50% pada pengujian kedua. Pada pengukuran tingkat presisi yaitu 85.76% pada pengujian pertama dan 84.21% pada pengujian kedua. Pada pengukuran tingkat recall yaitu 76% pada pengujian pertama dan 78.38% pengujian kedua. Pada skenario ini diperoleh nilai rata-rata akurasi 81.9643%, rata-rata presisi 84.9863% dan rata-rata recall 77.19%.



Gambar 4. Hasil pengukuran tingkat akurasi dan presisi pada cross fold validation k = 2
Sumber : Hasil Perhitungan

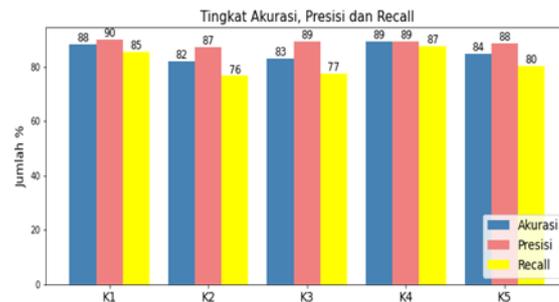
K = 4, Pada skenario ini yang dilakukan 4 kali pengujian dan diperoleh hasil pengukuran tingkat akurasi yaitu 87.14% lipatan pertama, 82.14% lipatan kedua, 82.86% lipatan ketiga dan 84.29% lipatan keempat. Pada pengukuran tingkat presisi yaitu 90.44% lipatan pertama, 87.47% lipatan kedua, 85.71% lipatan ketiga dan 85.32% lipatan keempat. Dari pengukuran tingkat recall diperoleh yaitu 83.09% pada lipatan pertama, 76.35% lipatan kedua, 78.46% lipatan ketiga dan 81.09% lipatan keempat. Pada skenario ini diperoleh nilai rata-rata akurasi 84.10%, rata-rata presisi 87.23% dan rata-rata recall 79.74%.



Gambar 5. Hasil pengukuran tingkat akurasi dan presisi pada cross fold validation k = 4

Sumber : Hasil Perhitungan

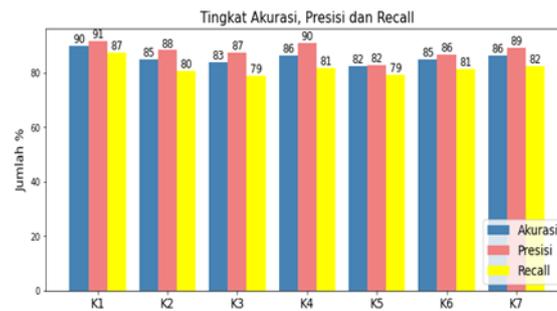
K = 5, Pada skenario ini yang dilakukan 5 kali pengujian dan diperoleh hasil pengukuran tingkat akurasi yaitu 88.39% lipatan pertama, 82.14% lipatan kedua, 83.04% lipatan ketiga, 89.29% lipatan keempat dan 84.82% lipatan kelima. Pada pengukuran tingkat presisi yaitu 90.01% lipatan pertama, 87.12% lipatan kedua, 89.33% lipatan ketiga, 89.33% lipatan keempat dan 88.74% lipatan kelima. Dari pengukuran tingkat recall diperoleh yaitu 85.48% pada lipatan pertama, 76.67% lipatan kedua, 77.38% lipatan ketiga, 87.62% lipatan keempat dan 80.24% lipatan kelima. Pada skenario ini diperoleh nilai rata-rata akurasi 85.53%, rata-rata presisi 88.90% dan rata-rata recall 81.47%.



Gambar 6. Hasil pengukuran tingkat akurasi dan presisi pada cross fold validation k = 5

Sumber : Hasil Perhitungan

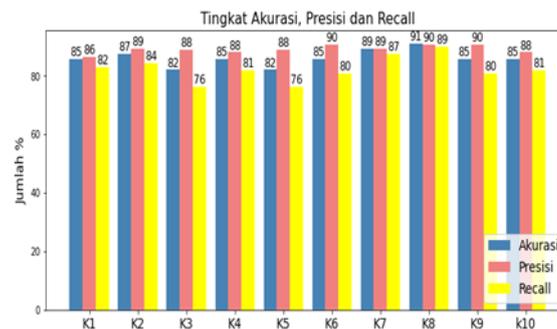
K = 7, Pada pengukuran tingkat presisi yaitu 91.67% lipatan pertama, 88.33% lipatan kedua, 87.53% lipatan ketiga, 90.98% lipatan keempat, 82.74% lipatan kelima, 86.83% lipatan keenam dan 89.14% lipatan ketujuh. Dari pengukuran tingkat recall diperoleh yaitu 87.33% pada lipatan pertama, 80.67% lipatan kedua, 79% lipatan ketiga, 81.67% lipatan keempat, 79.33% lipatan kelima, 81.33% lipatan keenam dan 82.33% lipatan ketujuh. Pada skenario ini diperoleh nilai rata-rata akurasi 85.53%, rata-rata presisi 88.17% dan rata-rata recall 81.66%.



Gambar 7. Hasil pengukuran tingkat akurasi dan presisi pada cross fold validation k = 7

Sumber : Hasil Perhitungan

K = 10, Pada pengukuran tingkat presisi yaitu 86.43% lipatan pertama, 89.38% lipatan kedua, 88.89% lipatan ketiga, 88.13% lipatan keempat, 88.89% lipatan kelima, 90.70% lipatan keenam, 89.33% lipatan ketujuh, 90.83% lipatan kedelapan, 90.70% lipatan kesembilan dan 88.13% lipatan kesepuluh. Dari pengukuran tingkat recall diperoleh yaitu 82.86% pada lipatan pertama, 84.29% lipatan kedua, 76.19% lipatan ketiga, 81.9% lipatan keempat, 76.19% lipatan kelima, 80.95% lipatan keenam, 87.62% lipatan ketujuh, 90% lipatan kedelapan, 80.95% lipatan kesembilan dan 81.9% lipatan kesepuluh. Pada skenario ini diperoleh nilai rata-rata akurasi 86.07%, rata-rata presisi 89.13% dan rata-rata 82.28%.



Gambar 8. Hasil pengukuran tingkat akurasi dan presisi pada cross fold validation k = 10

Sumber : Hasil Perhitungan

I. Hasil Analisis

Dari hasil skenario pengujian sentimen analisis pada komentar masyarakat terhadap produk Eiger melalui kolom komentar Instagram menggunakan metode Naive Bayes berikut ditampilkan rekapitulasi hasil pengukuran tingkat akurasi dan presisi.

2. Tingkat akurasi yang diperoleh Naive Bayes Classifier dalam mengklasifikasi data komentar masyarakat terhadap produk Eiger pada Instagram yaitu memiliki akurasi tertinggi sebesar 91.07% yang diperoleh pada skenario uji cross fold validation dengan k = 10 lipatan ke 8. Tingkat presisi yang diperoleh Naive Bayes Classifier dalam mengklasifikasi data komentar masyarakat terhadap produk Eiger pada Instagram yaitu memiliki akurasi tertinggi sebesar 91.67% yang diperoleh pada skenario uji cross fold validation dengan k = 7 lipatan ke 1. Sedangkan tingkat recall yang diperoleh Naive Bayes Classifier dalam mengklasifikasi data komentar masyarakat terhadap produk Eiger pada Instagram yaitu memiliki akurasi tertinggi sebesar 90% yang diperoleh pada skenario uji cross fold validation dengan k = 10 lipatan ke 8.
3. Pada skenario uji menggunakan Split data dengan pembagian data 20% data uji dan 80% data latih diperoleh hasil tingkat akurasi 85.71%, presisi 96.62% dan recall 83.49%.
4. Dari hasil analisis kemunculan kata yang kerap ditulis pada komentar produk Eiger yaitu antara lain mantap, beli, nyaman, keren, awet, bagus dan awet pada kluster sentimen positif. Sedangkan pada kluster negatif kemunculan kata yang mendominasi antara lain putus, mahal, jam, tali, pudar dan rusak.

DAFTAR PUSTAKA

- [1] Alimuddin, Budi. (2015). "Berebut Rezeki dari Ransel". Detiknews.com. Pada laman <http://news.detik.com/lapsus/3014890/berebut-rezeki-dari-ransel> diakses pada tanggal 10 Februari 2021.
- [2] Ellya. (2015). Promosi Wisata Lebih Cepat Lewat Sosmed dan Media Sosial. Pada laman <http://beritajateng.net/promosi-wisata-lebih-cepat-lewat-sosmed-dan-media-sosial/> diakses tanggal 24 Februari 2021.
- [3] Instagram. (2021). Akun Eiger @ eigerindostore. Pada laman <https://www.instagram.com/eigerindostore/> diakses pada tanggal 24 Februari 2021.
- [4] Nur, M. Y., & Santika, D. D. (2011). Analisis Sentimen Pada Dokumen Berbahasa Indonesia dengan Pendekatan Support Vector Machine. Konferensi Nasional Sistem dan Informatika, 9.
- [5] Routray, P., Swain, C. K., & Mishra, S. P. (2013). A survey on sentiment analysis. International Journal of Computer Applications, 76(10).
- [6] Zuhri, F., & Alamsyah, A. (2017). Analisis Sentimen Masyarakat Terhadap Brand Smartfren Menggunakan Naive Bayes Classifier di Forum Kaskus. eProceedings of Management, 4(1).